Tropospheric water vapour isotopologue data ($H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$) as obtained from NDACC/FTIR solar absorption spectra

Sabine Barthlott[1], Matthias Schneider[1], Frank Hase[1], Thomas Blumenstock[1], Matthäus Kiel[1],
Darko Dubravica[1], Omaira E. García[2], Eliezer Sepúlveda[2], Gizaw Mengistu Tsidu[3,4], Samuel Takele
Kenea[3,a], Michel Grutter[5], Eddy F. Plaza-Medina[5], Wolfgang Stremme[5], Kim Strong[6], Dan Weaver[6],
Mathias Palm[7], Thorsten Warneke[7], Justus Notholt[7], Emmanuel Mahieu[8], Christian Servais[8],
Nicholas Jones[9], David W. T. Griffith[9], Dan Smale[10], and John Robinson[10]

[1]Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology (KIT),
Karlsruhe, Germany
[2]Izaña Atmospheric Research Center, Agencia Estatal de Meteorología (AEMET),
Santa Cruz de Tenerife, Spain
[3]Department of Physics, Addis Ababa University, P.O. Box 1176, Addis Ababa, Ethiopia
[4]Botswana International University of Technology and Science (BIUST) Priv. Bag 16, Palapye, Botswana
[5]Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México, 04510 Mexico City, Mexico
[6]Department of Physics, University of Toronto, Toronto, Ontario, Canada
[7]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[8]Institute of Astrophysics and Geophysics, University of Liège, Liège, Belgium
[9]Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, New South Wales, Australia
[10]National Institute of Water and Atmospheric Research, Lauder, New Zealand
[a]now at: Department of Physics, Samara University, P.O. Box 132, Samara, Ethiopia

Correspondence to: Sabine Barthlott (sabine.barthlott@kit.edu)

Abstract. We report on the ground-based FTIR (Fourier transform infrared) tropospheric water vapour isotopologue remote sensing data that have been recently made available via the database of NDACC (Network for the Detection of Atmospheric Composition Change; ftp://ftp.cpc.ncep.noaa.gov/ndacc/MUSICA/) and via doi:10.5281/zenodo.48902. Currently, data are available for 12 globally distributed stations. They have been centrally retrieved and quality-filtered in the framework of the MUSICA project (MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water). We explain particularities of retrieving the water vapour isotopologue state (vertical distribution of $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$) and reveal the need for a new metadata template for archiving FTIR isotopologue data. We describe the format of different data components and give recommendations for correct data usage. Data are provided as two data types. The first type is best-suited for tropospheric water vapour distribution studies disregarding different isotopologues (comparison with radiosonde data, analyses of water vapour variability and trends, etc.). The second type is needed for analysing moisture pathways by means of {H_2O, δD}-pair distributions.

1   Introduction

Simultaneous observations of different tropospheric water isotopologues can provide valuable information on moisture source, transport, cloud processes, and precipitation (e.g. Dansgaard, 1964; Gat, 2000; Yoshimura et al., 2004). For identifying and analysing different tropospheric moisture pathways, the distribution of {$H_2O, \delta D$} pairs (e.g. Galewsky et al., 2005; Noone, 2012; González et al., 2016) or deuterium excess ($d = \delta D - 8\delta^{18}$; e.g. Pfahl and Sodemann, 2014; Steen-Larsen et al., 2014; Aemisegger et al., 2014) are particularly promising. The formula of deuterium excess is based on definitions used in studies such as Craig (1961a, b) and Dansgaard (1964). The δ notation is used to express the relation of the observed isotopologue ratio to the standard ratio VSMOW (Vienna Standard Mean Ocean Water), where $\delta D = \frac{HD^{16}O/H_2^{16}O}{VSMOW} - 1$ and $\delta^{18} = \frac{H_2^{18}O/H_2^{16}O}{VSMOW} - 1$.

In recent years, there has been significant progress in measuring the tropospheric water vapour isotopologues; remote sensing observations are particularly interesting since they can provide data for the free troposphere and they can be performed continuously (for cloud-free conditions). During MUSICA (MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water), a method has been developed to obtain tropospheric water vapour profiles as well as {$H_2O, \delta D$} pairs from ground-based FTIR (Fourier transform infrared) and space-based IASI (Infrared Atmospheric Sounding Interferometer) observations. The ground-based FTIR spectra are measured in the framework of the NDACC (Network for the Detection of Atmospheric Composition Change, www.ndacc.org). These spectra are of very high quality and have been available at some stations since the 1990s, thus offering long-term data records, which are of particular interest to climatological studies or for assessing the stability of long-term satellite data records. Network-wide consistent quality of the MUSICA NDACC/FTIR water vapour isotopologue data is ensured by central data processing and quality control (e.g. all data are quality-filtered by the XCO_2 method as presented in Barthlott et al., 2015). The high quality of the MUSICA ground- and space-based remote sensing data has been demonstrated by empirical validation studies (Schneider et al., 2016).

In this paper, we present the MUSICA NDACC/FTIR data as provided recently via the NDACC database (see also Barthlott et al., 2016). Our objective is to make a data user aware of the nature of the data and to give recommendations and explanations for correct data usage. Section 2 briefly describes the particularity of retrievals of the water vapour isotopologue state. It is shown that for achieving an optimal estimation of the tropospheric water vapour isotopologue state, we have to work with full state vectors consisting of different isotopologues, i.e. a state vector consisting of several trace gases. As a result, water vapour isotopologue data cannot be provided in exactly the same data format as other FTIR data. In Sect. 3 we describe the data format used for providing this new data via the NDACC database. Section 4 gives some insight into the data characteristics and recommendation for working with this new data set. In Sect. 6, we conclude and summarise our results. The summarised guidelines for correct data usage can be found in Table 3.

2   Optimal estimation of the water vapour isotopologue state

2.1   The principle of optimal estimation retrieval methods

Atmospheric remote sensing retrievals characterise an atmospheric state from a measured spectrum. However, such an inversion problem is often ill-posed (a lot of different atmospheric states can explain the measured spectrum). Consequently, for solving this problem, some kind of regularisation is required. This can be introduced by means of a cost function:

$$
\begin{aligned}
&[y - F(x, p)]^T S_\epsilon^{-1} [y - F(x, p)] \\
&+ [x - x_a]^T S_a^{-1} [x - x_a].
\end{aligned} \tag{1}
$$

Here, the first term is a measure of the difference between the measured spectrum (y) and the spectrum simulated for a given atmospheric state (x), where F represents the forward model, which simulates a spectrum y for a given state x, taking into account the actual measurement noise level (S_ϵ is the measurement noise covariance). Vector p represents auxiliary atmospheric parameters (like temperature) or instrumental characteristics (like the instrumental line shape). The second term of the cost function (1) is the regularisation term. It constrains the atmospheric solution state (x) towards an a priori most likely state (x_a), where kind and strength of the constraint are defined by the a priori covariance matrix S_a. The constrained solution is reached at the minimum of the cost function (1). This method of updating the knowledge about the a priori state with information from a measurement is known as the optimal estimation method, which is a standard remote sensing retrieval method (see Rodgers, 2000, for more details) in which validity of the optimal estimation solution strongly depends on a correct and comprehensive description of the a priori state by means of x_a and S_a. For readers that are not familiar with remote sensing mathematics, we added some mathematical foundations in Appendix A.

2.2   Correct description of the water vapour isotopologue state

Water vapour isotopologues with reasonably strong and well-discernible spectral infrared signatures are $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$. Hereafter, we will refer to the combined states of $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ as the full water vapour isotopologue state.

Tropospheric water vapour shows a strong variation (in space and time) and it can be better described by a log-normal than by a normal distribution (Hase et al., 2004; Schneider et al., 2006). Furthermore, the different isotopologues vary mostly in equal measure, i.e. their variations are strongly correlated and the variation in the isotopologue ratios is much smaller. For instance, the variation in $\ln\left[\mathrm{HD^{16}O}\right] - \ln\left[\mathrm{H_2^{16}O}\right]$ is about 1 order of magnitude smaller and the variations in $\ln\left[\mathrm{H_2^{18}O}\right] - \ln\left[\mathrm{H_2^{16}O}\right]$ is about 2 orders of magnitude smaller than the correlated variations in $\ln\left[\mathrm{H_2^{16}O}\right]$, $\ln\left[\mathrm{H_2^{18}O}\right]$, and $\ln\left[\mathrm{HD^{16}O}\right]$ (Craig and Gordon, 1965). An elegant method for correctly describing the nature of the full water vapour isotopologue state is to work on a logarithmic scale (due to the log-normal distribution characteristic) and with the following three states:

– the humidity-proxy state:

$$\mathrm{humidity} = \frac{1}{3}\left(\ln\left[\mathrm{H_2^{16}O}\right] + \ln\left[\mathrm{H_2^{18}O}\right] + \ln\left[\mathrm{HD^{16}O}\right]\right); \quad (2)$$

– the δD-proxy state:

$$\delta\mathrm{D} = \ln\left[\mathrm{HD^{16}O}\right] - \ln\left[\mathrm{H_2^{16}O}\right]; \quad (3)$$

– the deuterium-excess proxy state (d-proxy state):

$$d = \ln\left[\mathrm{HD^{16}O}\right] - \ln\left[\mathrm{H_2^{16}O}\right] \\ - 8\left(\ln\left[\mathrm{H_2^{18}O}\right] - \ln\left[\mathrm{H_2^{16}O}\right]\right). \quad (4)$$

The water vapour isotopologue state can be expressed on the basis of $\left\{\ln\left[\mathrm{H_2^{16}O}\right], \ln\left[\mathrm{H_2^{18}O}\right], \ln\left[\mathrm{HD^{16}O}\right]\right\}$ or on the basis of the proxies of $\{\mathrm{humidity}, \delta\mathrm{D}, d\}$. Both expressions are equivalent. Each basis has the dimension $nol \times 3$, where nol is the number of levels of the radiative transfer model atmosphere and three different combinations of the three different isotopologues have to be considered. In the following, the full water vapour isotopologue state vector expressed on the $\left\{\ln\left[\mathrm{H_2^{16}O}\right], \ln\left[\mathrm{H_2^{18}O}\right], \ln\left[\mathrm{HD^{16}O}\right]\right\}$ basis and the $\{\mathrm{humidity}, \delta\mathrm{D}, d\}$-proxy basis will be referred to as $\boldsymbol{x}^1$ and $\boldsymbol{x}^{1\prime}$, respectively, where index "l" stands for logarithmic scale. A basis transformation can be achieved by operator $\mathbf{P}$:

$$\mathbf{P} = \begin{pmatrix} \frac{1}{3}\mathbb{1} & \frac{1}{3}\mathbb{1} & \frac{1}{3}\mathbb{1} \\ -\mathbb{1} & 0 & \mathbb{1} \\ 7\mathbb{1} & -8\mathbb{1} & \mathbb{1} \end{pmatrix}. \quad (5)$$

Here, the nine matrix blocks have the dimension $nol \times nol$, $\mathbb{1}$ stands for an identity matrix, and the state vectors $\boldsymbol{x}^1$ and $\boldsymbol{x}^{1\prime}$ are related by

$$\boldsymbol{x}^{1\prime} = \mathbf{P}\boldsymbol{x}^1. \quad (6)$$

Similarly, covariance matrices can be expressed in the two basis systems and the respective matrices $\mathbf{S}^l$, and $\mathbf{S}^{l\prime}$ are related by

$$\mathbf{S}^{l\prime} = \mathbf{P}\mathbf{S}^l\mathbf{P}^T. \quad (7)$$

The variation in humidity, δD, and deuterium excess have different magnitudes. This can be well considered by giving the a priori covariance matrix in the $\{\mathrm{humidity}, \delta\mathrm{D}, d\}$-proxy basis:

$$\mathbf{S}_a^{l\prime} = \begin{pmatrix} \mathbf{S}_{a_{\mathrm{hum}}}^l & 0 & 0 \\ 0 & \mathbf{S}_{a_{\delta\mathrm{D}}}^l & 0 \\ 0 & 0 & \mathbf{S}_{a_d}^l \end{pmatrix}. \quad (8)$$

This assumes that humidity, δD, and deuterium excess vary independently, which is actually not the case in the lower/middle troposphere. However, the correlation between humidity, δD, and deuterium excess is a detail. It is important that the different magnitudes of variability are taken into account (the entries of $\mathbf{S}_{a_{\mathrm{hum}}}^l$ are much larger than the entries of $\mathbf{S}_{a_{\delta\mathrm{D}}}^l$ and $\mathbf{S}_{a_d}^l$).

The covariance matrix $\mathbf{S}_a^l$ can be calculated as (according to Eq. 7)

$$\mathbf{S}_a^l = \mathbf{P}^{-1}\mathbf{S}_a^{l\prime}\mathbf{P}^{-T}$$

$$= \begin{pmatrix} \mathbb{1} & -\frac{3}{8}\mathbb{1} & \frac{1}{24}\mathbb{1} \\ \mathbb{1} & -\frac{1}{4}\mathbb{1} & -\frac{1}{12}\mathbb{1} \\ \mathbb{1} & \frac{5}{8}\mathbb{1} & \frac{1}{24}\mathbb{1} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{a_{\mathrm{hum}}}^l & 0 & 0 \\ 0 & \mathbf{S}_{a_{\delta\mathrm{D}}}^l & 0 \\ 0 & 0 & \mathbf{S}_{a_d}^l \end{pmatrix}$$

$$\begin{pmatrix} \mathbb{1} & \mathbb{1} & \mathbb{1} \\ -\frac{3}{8}\mathbb{1} & -\frac{1}{4}\mathbb{1} & \frac{5}{8}\mathbb{1} \\ \frac{1}{24}\mathbb{1} & -\frac{1}{12}\mathbb{1} & \frac{1}{24}\mathbb{1} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{S}_{a_{\mathrm{hum}}}^l + \frac{9}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l + \frac{1}{24^2}\mathbf{S}_{a_d}^l & \mathbf{S}_{a_{\mathrm{hum}}}^l + \frac{6}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l - \frac{2}{24^2}\mathbf{S}_{a_d}^l \\ \mathbf{S}_{a_{\mathrm{hum}}}^l + \frac{6}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l - \frac{2}{24^2}\mathbf{S}_{a_d}^l & \mathbf{S}_{a_{\mathrm{hum}}}^l + \frac{4}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l + \frac{4}{24^2}\mathbf{S}_{a_d}^l \\ \mathbf{S}_{a_{\mathrm{hum}}}^l - \frac{15}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l + \frac{1}{24^2}\mathbf{S}_{a_d}^l & \mathbf{S}_{a_{\mathrm{hum}}}^l - \frac{10}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l - \frac{2}{24^2}\mathbf{S}_{a_d}^l \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{S}_{a_{\mathrm{hum}}}^l - \frac{15}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l + \frac{1}{24^2}\mathbf{S}_{a_d}^l \\ \mathbf{S}_{a_{\mathrm{hum}}}^l - \frac{10}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l - \frac{2}{24^2}\mathbf{S}_{a_d}^l \\ \mathbf{S}_{a_{\mathrm{hum}}}^l + \frac{25}{8^2}\mathbf{S}_{a_{\delta\mathrm{D}}}^l + \frac{1}{24^2}\mathbf{S}_{a_d}^l \end{pmatrix}. \quad (9)$$

This matrix $\mathbf{S}_a^l$ correctly captures the covariance of the a priori state in the $\left\{\ln\left[\mathrm{H_2^{16}O}\right], \ln\left[\mathrm{H_2^{18}O}\right], \ln\left[\mathrm{HD^{16}O}\right]\right\}$ basis and it is important for a correct formulation of the cost function (1) and thus for setting up a correctly constrained optimal estimation retrieval of the water vapour isotopologue state.

Matrix $\mathbf{S}_a^l$ reveals the complex covariances between the state vector components of $\boldsymbol{x}^1$, in which it has to be considered that the entries of $\mathbf{S}_{a_{\mathrm{hum}}}^l$ are much larger than the entries of $\mathbf{S}_{a_{\delta\mathrm{D}}}^l$ and significantly larger than the entries of $\mathbf{S}_{a_d}^l$, which means that the nine $nol \times nol$ blocks of $\mathbf{S}_a^l$ have almost the same entries. The variations in $\ln\left[\mathrm{H_2^{16}O}\right]$, $\ln\left[\mathrm{H_2^{18}O}\right]$, and $\ln\left[\mathrm{HD^{16}O}\right]$ are strongly correlated and we cannot work with individual state vectors that report the states of the different isotopologues independently. Instead, we have to describe the full isotopologue state by a single state vector with the dimension $nol \times 3$. This is the reason why MUSICA NDACC/FTIR water vapour isotopologue data cannot be provided in the same data format as other NDACC/FTIR

data. A slightly extended data format is needed, which is described in the next section.

2.3 The MUSICA (v2015) ground-based NDACC/FTIR retrieval setup

In Schneider et al. (2012), the MUSICA NDACC/FTIR retrieval setup, such as interfering gases and temperature fit, is described in great detail. Here, we focus on the modification made for the retrieval version (v2015).

For the previous retrieval version, we used 11 spectral windows with lines of water vapour isotopologues (see Fig. 2 of Schneider et al., 2012). For the final MUSICA retrieval version (v2015), we removed windows with strong $H_2^{16}O$ lines and replaced them by windows with weaker lines. By this modification, we want to make sure that even for very humid sites observed spectral lines do not saturate. Furthermore, we add a second window where a $H_2^{18}O$ signature dominates. The nine spectral water vapour isotopologue windows are depicted in Fig. 1 for an observation and fit, which are typical for the different NDACC stations. In addition, we fit the three spectral windows with the CO_2 lines, which is beneficial for atmospheric temperature retrievals (the CO_2 lines are the same as for the previous retrieval version: between 2610.35 and 2610.8, 2613.7 and 2615.4, and 2626.3 and 2627.0 cm^{-1}; not shown). All of these spectral windows are observed within NDACC filter #3.

For v2015, we perform an optimal estimation of the {humidity, δD, d}-proxy states as explained in Sect. 2.2. This is a further development of the previous retrieval version (there the optimal estimation was made for the {humidity, δD}-proxy states; Sects. 3 and 4 of Schneider et al., 2012). Consistent with the previous version, we perform simultaneous but individual fits (no cross-constraints) for profiles of the water vapour isotopologue $H_2^{17}O$, temperature, and interfering species CO_2, O_3, N_2O, CH_4, and HCl.

For the previous version, we used HITRAN 2008 line parameters (Rothman et al., 2009), whereas for v2015 we work with HITRAN 2012 parameters (Rothman et al., 2013). For both versions, the water vapour isotopologues parameters have been adjusted for the speed-dependent Voigt line shape (Schneider et al., 2011). For further optimisation of the HITRAN parameters, we used high-quality H_2O and δD aircraft in situ profile references, coincident FTIR spectra, and FTIR spectra measured at three rather distinct sites (Izaña on Tenerife, subtropical ocean; Karlsruhe, central Europe; and Kiruna, northern Sweden). This method for line parameter optimisation by means of atmospheric spectra is described in detail in Schneider and Hase (2009) and Schneider et al. (2011). We changed line intensities and broadening parameters by about 5–10 %, which is in agreement with the uncertainty values as given in the HITRAN parameter files. More details on the modification of the line parameters as shown in Fig. 1 are given in Appendix B.

The empirical assessment study of Schneider et al. (2016) suggests an accuracy for the MUSICA (v2015) NDACC/FTIR H_2O and δD products of about 10 % and 10 ‰, respectively.

3   Data set description

MUSICA NDACC/FTIR water vapour isotopologue data are available via the NDACC database and can be accessed via ftp://ftp.cpc.ncep.noaa.gov/ndacc/MUSICA/ and via doi:10.5281/zenodo.48902. The data are provided as HDF4 files and they have been generated in compliance with GEOMS (Generic Earth Observation Metadata Standard).

For isotopologue data, a new metadata template, GEOMS-TE-FTIR-ISO-001, has been set up. It is almost identical to the GEOMS-TE-FTIR template (used for all other FTIR data provided via NDACC in HDF4 format) but has the additional variable "CROSSCORRELATE.N". For MUSICA data, this new variable has three entries: "$H_2^{16}O$", "$H_2^{18}O$", and "$HD^{16}O$". This enables the three trace gases $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ to be provided as one full state vector (vector with $nol \times 3$ elements) together with their full averaging kernels and full error covariances (both of $(nol \times 3) \times (nol \times 3)$ dimension) in one data file. To be compliant with GEOMS, the data are provided on a linear scale (and not on a logarithmic scale on which the inversion is performed; see Sect. 2).

All the isotopologue data have been normalised with respect to their natural isotopologue abundances. These natural abundances are 0.997317 for $H_2^{16}O$, 0.002000 for $H_2^{18}O$, and 3.106930×10^{-4} for $HD^{16}O$. This normalisation means that δD can directly be calculated as $\left(\frac{[HD^{16}O]}{[H_2^{16}O]} - 1 \right)$ from provided $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ amounts. Similarly, deuterium excess can be calculated as $\left(\frac{[HD^{16}O]}{[H_2^{16}O]} - 1 \right) - 8 \times \left(\frac{[H_2^{18}O]}{[H_2^{16}O]} - 1 \right)$.

Two different data types are provided. The first type is stored in HDF files called "ftir.iso.h2o". These files report the best estimate of the $H_2^{16}O$ state, but its usefulness for isotopologue studies (e.g. in the context of {H_2O, δD}-distribution plots) is significantly compromised. The second type is stored in HDF files called "ftir.iso.post.h2o". This data type should be used for analysing {H_2O, δD}-distribution plots. Section 4 gives more details on the different data types.

The number of stations contributing to the MUSICA NDACC/FTIR data set is gradually increasing and Table 1 gives an overview of the current status (status for January 2016) together with mean DOFS (degrees of freedom of signal) for the two data types. The stations are well distributed from the Arctic to the Antarctic and in some cases offer data from the late 1990s onward. A further extension of this data set to other sites or for some stations to measurements made
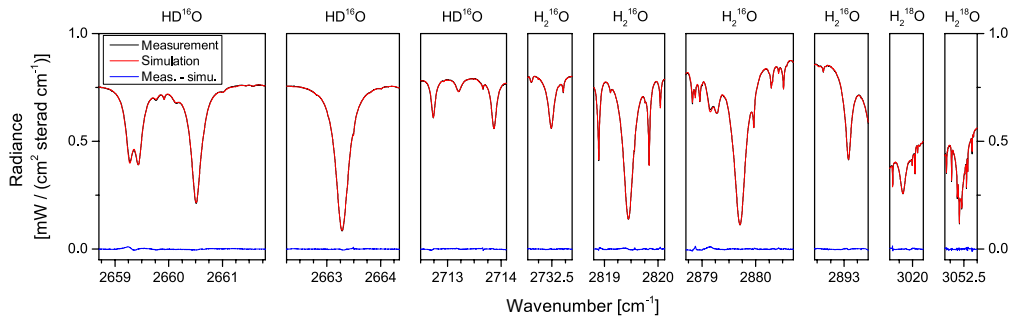
Figure 1. The spectral windows used for the MUSICA ground-based NDACC/FTIR retrievals. Shown is an example of a typical measurement (Karlsruhe, 15 September 2011, 12:03 UT; solar elevation: 43.1°; H_2O slant column: 22.9 mm). Black line: measurement; red line: simulation; blue line: residual (difference between measurement and simulation).

Table 1. List of current MUSICA NDACC/FTIR sites (ordered from north to south) and available MUSICA data record. DOFS (type 1) reports the typical trace of $\mathbf{A}^1_{H_2O}$ and DOFS (type 2) that of $\mathbf{A}^{1*}_{11} \approx \mathbf{A}^{1*}_{22}$ for optimal estimation of $\{H_2O, \delta D\}$ pairs (example kernel plotted in Fig. 5).

Site	Location	Altitude	Data record	DOFS (type 1)	DOFS (type 2)
Eureka, Canada	80.1° N, 86.4° W	610 m a.s.l.	2006–2014	2.9	1.7
Ny-Ålesund, Norway	78.9° N, 11.9° E	21 m a.s.l.	2005–2014	2.8	1.6
Kiruna, Sweden	67.8° N, 20.4° E	419 m a.s.l.	1996–2014	2.8	1.6
Bremen, Germany	53.1° N, 8.9° E	27 m a.s.l.	2004–2014	2.8	1.6
Karlsruhe, Germany	49.1° N, 8.4° E	110 m a.s.l.	2010–2014	2.8	1.6
Jungfraujoch, Switzerland	46.6° N, 8.0° E	3580 m a.s.l.	1996–2014	2.7	1.6
Izaña, Tenerife, Spain	28.3° N, 16.5° W	2367 m a.s.l.	1999–2014	2.9	1.7
Altzomoni, Mexico	19.1° N, 98.7° W	3985 m a.s.l.	2012–2014	2.7	1.7
Addis Ababa, Ethiopia	9.0° N, 38.8° E	2443 m a.s.l.	2009–2013	2.6	1.6
Wollongong, Australia	34.5° S, 150.9° E	30 m a.s.l.	2007–2014	2.7	1.6
Lauder, New Zealand	45.1° S, 169.7° E	370 m a.s.l.	1997–2014	2.8	1.6
Arrival Heights, Antarctica	77.8° S, 166.7° E	250 m a.s.l.	2002–2014	2.7	1.4

in the beginning of the 1990s is feasible but was not possible within the MUSICA project.

3.1 Averaging kernels

An averaging kernel describes how a retrieved state vector responds to variations in the real atmospheric state vector. The MUSICA NDACC/FTIR isotopologue state vector consists of three trace gases and the corresponding full averaging kernel matrix $\mathbf{A}$ consists of nine blocks, each of which is a *nol* × *nol* matrix. In total, $\mathbf{A}$ has the dimension $(nol \times 3) \times (nol \times 3)$:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} \end{pmatrix}. \tag{10}$$

The three blocks along the diagonal describe the direct responses: block $\mathbf{A}_{11}$ for the response of the $H_2^{16}O$ retrieval product to real atmospheric $H_2^{16}O$ variations, block $\mathbf{A}_{22}$ for the response of the $H_2^{18}O$ retrieval product to real atmospheric $H_2^{18}O$ variations, and block $\mathbf{A}_{33}$ for the response

of the $HD^{16}O$ retrieval product to real atmospheric $HD^{16}O$ variations. The outer diagonal blocks describe the cross responses: two blocks for the response of the $H_2^{16}O$ retrieval product to real atmospheric $H_2^{18}O$ and $HD^{16}O$ variations ($\mathbf{A}_{12}$ and $\mathbf{A}_{13}$, respectively), two blocks for the response of the $H_2^{18}O$ retrieval product to real atmospheric $H_2^{16}O$ and $HD^{16}O$ variations ($\mathbf{A}_{21}$ and $\mathbf{A}_{23}$, respectively), and two blocks for the response of the $HD^{16}O$ retrieval product to real atmospheric $H_2^{16}O$ and $H_2^{18}O$ variations ($\mathbf{A}_{31}$ and $\mathbf{A}_{32}$, respectively). The different blocks of the matrix (see Fig. 2) can be easily identified by the values of the variable "CROSS-CORRELATE.N": 1 = $H_2^{16}O$, 2 = $H_2^{18}O$, and 3 = $HD^{16}O$.

3.2 Error covariances

For error calculations, the same uncertainty sources are assumed for all MUSICA NDACC/FTIR stations and are grouped into statistical and systematic errors. An overview of the uncertainty assumptions is given in Table 2. Error covariances for each of these uncertainty sources are calculated according to Rodgers (2000). Total error covariances are ob-
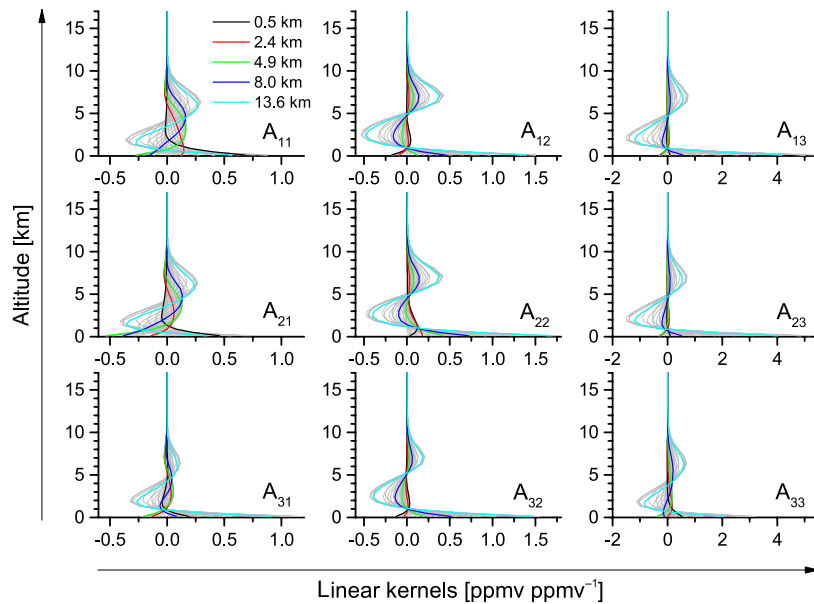
Figure 2. Column entries of the nine blocks of the full averaging kernel matrix **A** (see Eq. 10). Shown are some columns of the kernel matrix for 0.5 km (black), 2.4 km (red), 4.9 km (green), 8.0 km (blue), and 13.6 km (light blue). This kernel is for the retrieval with the spectrum and fit as shown in Fig. 1.

tained as the sum of the individual error covariances, calculated separately for statistical and systematic errors. Full error covariance matrices have the same matrix dimension as the full averaging kernel matrix and provide information about the errors and how the errors between different altitudes as well as between the three different isotopologues are correlated.

3.3   Column densities

Partial and total column densities are also provided as well as column sensitivity averaging kernels and column error covariances. Partial columns are calculated for the layers between the nol atmospheric levels. A full water vapour isotopologue partial column state vector has $(nol - 1) \times 3$ elements (i.e. $nol - 1$ elements for each of the three isotopologues). The column sensitivity averaging kernel has the dimension $((nol - 1) \times 3) \times 3$ and describes the sensitivities of the total column retrieval product (columns of $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$) with respect to real atmospheric variations in the isotopologues' partial columns. Column error covariances are provided for statistical and systematic errors independently and are matrices with the dimension 3×3. They describe total column errors for the three isotopologues and how these errors are correlated.

All column data are calculated from mixing ratio data (mixing ratio state vector, error covariances, and averaging kernels). For a conversion of mixing ratios (ppmv) to number densities (molec. cm^{-3}), we use the ideal gas law and

work with retrieved temperature profiles. Temperature and pressure profiles are also provided in the HDF files.

Please note: since GEOMS requires the same number of elements of column data and the respective mixing ratio data, value 0 has been added for all column data values at level nol.

4   Recommendations for data usage

In order to be compliant with GEOMS, data are provided for the full state vector consisting of the $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ state vectors and in mixing ratios (ppmv). However, actually the MUSICA NDACC/FTIR retrieval works on a logarithmic scale and performs an optimal estimation of combined isotopologue states (see Sect. 2). These details have to be considered when working with these data.

4.1   Transfer to a logarithmic scale

For operations with averaging kernels (e.g. when adjusting model data to the sensitivity of a remote sensing system), it has to be considered that the retrieval works on a logarithmic scale, because only on this scale are linearity assumptions valid. Therefore, it is strongly recommended to transfer the averaging kernels to a logarithmic scale; in doing so, it has to be considered that the derivatives are calculated for the state as given by the retrieval state vector x. The full retrieval state vector consists of the retrieved $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ states, i.e. it is a vector with $nol \times 3$ elements x_i (with i between 1 and $nol \times 3$). The full averaging kernel matrix has the dimension $(nol \times 3) \times (nol \times 3)$. Figure 2 plots the

Table 2. Uncertainty sources used for the error estimation. The second column gives the assumed uncertainty value and the third column the assumed partitioning between statistical and systematic sources.

Error source	Uncertainty	Statistical/systematic
Measurement noise	0.4 %	100/0
Baseline (channelling, assuming 4 frequencies: 0.005, 0.2, 1.0, and 3.0 cm^{-1})	0.2 %	50/50
Baseline (offset)	0.1 %	50/50
Instrumental line shape (modulation efficiency and phase error)	10 % and 0.1 rad	50/50
Temperature profile	3 K (surface–5 km a.s.l.)	70/30
	3 K (5–12 km a.s.l.)	70/30
	3 K (above 12 km a.s.l.)	70/30
Line of sight	0.1°	90/10
Solar lines (intensity and ν scale)	1 % and 10^{-6}	80/20
Spectroscopic parameters (S and γ)	+1 % ($H_2^{16}O$)	0/100
	−1 % ($H_2^{18}O$)	0/100
	+2 % ($HD^{16}O$)	0/100

full averaging kernel matrix provided on a linear scale (**A**). It has the entries $a_{i, j}$ (i and j are between 1 and $nol \times 3$, where i is the row index and j the column index). Entry $a_{i, j}$ is a derivative, describing how the retrieved state vector element i responds to a variation in the real atmospheric state vector element j. Since $\partial \ln x = \frac{\partial x}{x}$, entry $a_{i, j}$ has to be modified for operations on a logarithmic scale as follows:

$$a_{i, j}^l = a_{i, j} \frac{x_j}{x_i}. \tag{11}$$

Here, $a_{i, j}^l$ are the entries of the full averaging kernel matrix transferred to a logarithmic scale ($\mathbf{A}^l$), which is plotted in Fig. 3. This matrix $\mathbf{A}^l$ is the averaging kernel matrix for a scale on which the linearity assumptions are valid and on which kernel operations should be performed.

For the purpose of error analyses, it is also very useful to transfer the error covariances on a logarithmic scale. On this scale, the δD and deuterium-excess error covariances can be made available in an elegant manner (see Sect. 4.4). The covariance matrix entry $s_{i, j}$ is the covariance between the state vector elements i and j (the value of these elements is x_i and x_j, respectively). On a logarithmic scale, the covariance matrix entry is

$$s_{i, j}^l = \frac{s_{i, j}}{x_i x_j}. \tag{12}$$

The covariance matrices on a logarithmic scale ($\mathbf{S}^l$) are composed of the entries $s_{i, j}^l$.

4.2 Comparison studies

For many purposes, remote sensing data may be compared to other data. For instance, they may be compared to vertically resolved observational data in order to empirically assess the quality of the different data sets, or they are compared to model data in order to investigate model performances. For such comparisons, it is important to consider the sensitivity of the remote sensing system. While highly resolved profile data or model data generally capture atmospheric signals well even on rather small scales, remote sensing data report atmospheric signals according to the averaging kernel. In order to make different data sets comparable, we have to convolve the data with the averaging kernels. A full water vapour isotopologue state vector obtained from highly resolved profile measurements or model calculations ($\boldsymbol{x}_{\text{data}}^l$) will be observed by a remote sensing system as $\boldsymbol{x}_{\text{data, conv}}^l$:

$$\boldsymbol{x}_{\text{data, conv}}^l = \mathbf{A}^l \left(\boldsymbol{x}_{\text{data}}^l - \boldsymbol{x}_{\text{a}}^l\right) + \boldsymbol{x}_{\text{a}}^l. \tag{13}$$

Here, $\boldsymbol{x}_{\text{a}}^l$ means the full water vapour isotopologue a priori state vector on a logarithmic scale.

4.3 Simple water vapour analyses (neglecting the different isotopologues)

About 99.7 % of all water vapour is present in form of the isotopologue $H_2^{16}O$, and for studies in which the different isotopologues are not distinguished, one should work with the retrieved $H_2^{16}O$ data. One example is comparison of the MUSICA NDACC/FTIR remote sensing data with water vapour data measured in situ by radiosondes. Since radiosondes do not distinguish between isotopologues, they report almost exclusively $H_2^{16}O$ concentrations and amounts.

In order to simulate how a remote sensing system observes a $\ln\left[H_2^{16}O\right]$ radiosonde profile, we have to convolve the state vector determined from radiosonde data ($\boldsymbol{x}_{\text{RS}}^l$) us-

www.earth-syst-sci-data.net/9/15/2017/
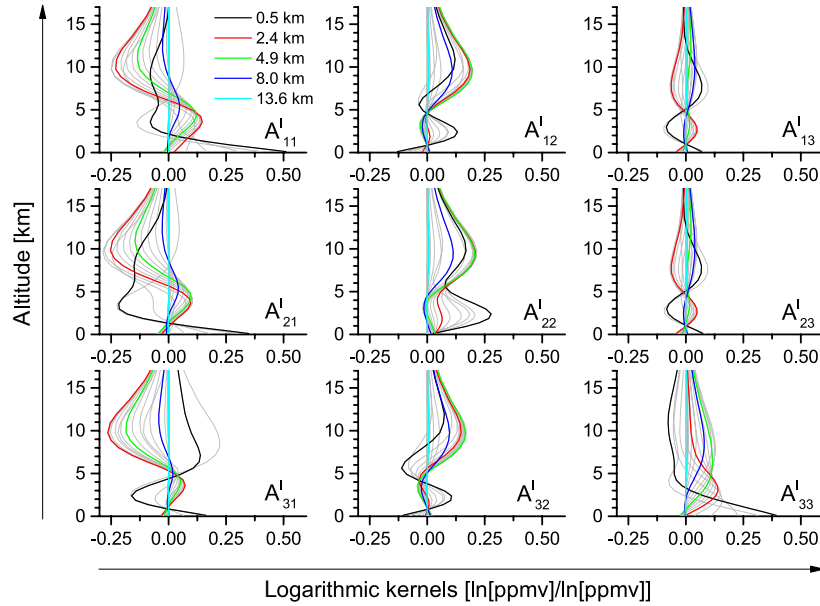
Earth Syst. Sci. Data, 9, 15–29, 2017

Figure 3. Same as Fig. 2 but for the logarithmic-scale kernel matrix $\mathbf{A}^l$.

ing the averaging kernel blocks $\mathbf{A}^l_{11}$, $\mathbf{A}^l_{12}$, and $\mathbf{A}^l_{13}$ (these blocks are depicted in Fig. 3). The radiosonde only measures $\ln\left[H_2^{16}O\right]$; it does not measure $\ln\left[H_2^{18}O\right]$ and $\ln\left[HD^{16}O\right]$. However, $\ln\left[H_2^{18}O\right]$ and $\ln\left[HD^{16}O\right]$ are strongly correlated with $\ln\left[H_2^{16}O\right]$:

$$\ln\left[H_2^{18}O\right] - \ln\left[H_2^{18}O\right]_a = \ln\left[H_2^{16}O\right] - \ln\left[H_2^{16}O\right]_a$$
$$+ \frac{1}{8}\{\delta D - \delta D_a - (d - d_a)\}, \tag{14}$$

and

$$\ln\left[HD^{16}O\right] - \ln\left[HD^{16}O\right]_a = \ln\left[H_2^{16}O\right] - \ln\left[H_2^{16}O\right]_a$$
$$+ \{\delta D - \delta D_a\} \tag{15}$$

(see Eqs. 3 and 4). Here, index a identifies a priori data. Like all full isotopologue state vectors, $\boldsymbol{x}^l_{RS}$ has $nol \times 3$ components. The first section of nol components is determined by the in situ measured $\ln\left[H_2^{16}O\right]$ profile. The second section of nol components is also determined by the in situ measured $\ln\left[H_2^{16}O\right]$ profile, but we have to consider an uncertainty of these vector components according to the uncertainty covariance matrix $\frac{1}{8^2}(\mathbf{S}^l_{a_{\delta D}} + \mathbf{S}^l_{a_d})$ (see Eq. 14). The uncertainty is due to the fact that a radiosonde does not measure $H_2^{18}O$. The third section of nol components is similarly determined by the in situ measured $\ln\left[H_2^{16}O\right]$ profile, but with an uncertainty according to the uncertainty covariance matrix $\mathbf{S}^l_{a_{\delta D}}$ (see Eq. 15), in which the uncertainty is caused by missing $HD^{16}O$ radiosonde measurements. Hence, a $\ln\left[H_2^{16}O\right]$ radiosonde profile is equivalent to a $\ln\left[H_2^{16}O\right]$ remote sens-

ing observation of

$$\left(\mathbf{A}^l_{11} + \mathbf{A}^l_{12} + \mathbf{A}^l_{13}\right)\left(\boldsymbol{x}^l_{RS} - \boldsymbol{x}^l_a\right) + \boldsymbol{x}^l_a, \tag{16}$$

in which there is an uncertainty in the equivalency of

$$\frac{1}{8^2}\mathbf{A}^l_{12}\left(\mathbf{S}^l_{a_{\delta D}} + \mathbf{S}^l_{a_d}\right)\mathbf{A}^{l\,T}_{12} + \mathbf{A}^l_{13}\mathbf{S}^l_{a_{\delta D}}\mathbf{A}^{l\,T}_{13}, \tag{17}$$

which is due to missing radiosonde observations of $H_2^{18}O$ and $HD^{16}O$.

Equation 16 reveals that the $nol \times nol$ averaging kernel matrix $\mathbf{A}^l_{H_2O} = \mathbf{A}^l_{11} + \mathbf{A}^l_{12} + \mathbf{A}^l_{13}$ is a good proxy for the remote sensing system's sensitivity with respect to $H_2^{16}O$ and thus water vapour in general (about 99.7 % of all water vapour molecules are $H_2^{16}O$ isotopologues). The columns of the matrix $\mathbf{A}^l_{H_2O}$ are plotted in Fig. 4.

4.4 Utility of the {humidity, δD, d}-proxy basis and data a posteriori processing

The MUSICA NDACC/FTIR retrieval performs an optimal estimation of the humidity, δD, and deuterium-excess proxy states. Although optimal estimation of these states is made in a single retrieval process, it is made for each of the three states independently, meaning, for instance, that the estimation is not optimal for the {H_2O, δD} pairs. The reason is that the remote sensing system's sensitivity for the humidity state is much higher than for the δD-state (and significantly higher than for the deuterium-excess state). The problem becomes clearly visible by transforming the full averaging kernel matrix onto the {humidity, δD, d}-proxy basis:

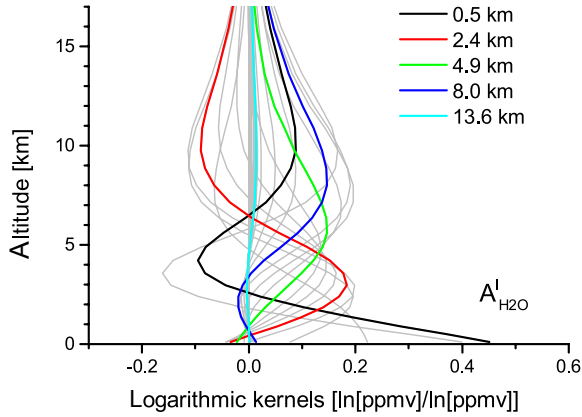$$\mathbf{A}^{l'} = \mathbf{P}\mathbf{A}^l\mathbf{P}^{-1}. \tag{18}$$

Figure 4. Logarithmic-scale kernel matrix $\mathbf{A}^1_{H_2O} = \mathbf{A}^1_{11} + \mathbf{A}^1_{12} + \mathbf{A}^1_{13}$, with the $\mathbf{A}^1$ blocks as plotted in Fig. 3.

Here, $\mathbf{A}^{1'}$ is the full averaging kernel matrix in the $\{humidity, \delta D, d\}$-proxy basis. It is depicted in Fig. 5, which clearly reveals larger sensitivity for humidity (kernel block $\mathbf{A}^1_{11}{}'$) than for δD (kernel block $\mathbf{A}^1_{22}{}'$). Data with this characteristic cannot be used in the context of $\{H_2O, \delta D\}$-pair distribution analyses.

A transformation of the error covariance matrices onto the $\{humidity, \delta D, d\}$-proxy basis (for transformation operation, see Eq. 7) is very helpful for analysing the error characteristics of the H_2O, δD, and deuterium-excess data because error covariances expressed in the $\{humidity, \delta D, d\}$-proxy basis are good proxies for the error covariances of H_2O, δD, and deuterium excess (for more details see discussion in Sect. 4.2 of Schneider et al., 2012).

4.4.1 A posteriori processing for a quasi-optimal estimation of $\{H_2O, \delta D\}$ pairs

During MUSICA, an a posteriori processing method for obtaining a quasi-optimal estimation product of $\{H_2O, \delta D\}$ pairs has been developed. The a posteriori processing brings about a moderate reduction in the H_2O sensitivity and in the δD cross dependencies on H_2O (Sect. 4.2 in Schneider et al., 2012). The operation has to be performed with logarithmic-scale full state vector, averaging kernel matrix, and covariance matrices (x^1, $\mathbf{A}^1$ and $\mathbf{S}^1$):

$$x^{1*} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}(x^1 - x^1_a) + x^1_a, \tag{19}$$

$$\mathbf{A}^{1*} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}\mathbf{A}^1, \tag{20}$$

and

$$\mathbf{S}^{1*} = \mathbf{P}^{-1}\mathbf{C}\mathbf{P}\mathbf{S}^1\mathbf{P}^T\mathbf{C}^T\mathbf{P}^{-T}. \tag{21}$$

Here, x^{1*}, $\mathbf{A}^{1*}$, and $\mathbf{S}^{1*}$ are a posteriori-processed logarithmic-scale full state vector, averaging kernel matrix, and covariance matrices, respectively. Operator $\mathbf{P}$ is

introduced by Eq. (5) and the a posteriori operator $\mathbf{C}$ is

$$\mathbf{C} = \begin{pmatrix} \mathbf{A}^1_{22}{}' & 0 & 0 \\ -\mathbf{A}^1_{21}{}' & \mathbb{1} & 0 \\ 0 & 0 & \mathbb{1} \end{pmatrix}. \tag{22}$$

Figure 6 depicts the a posteriori processed averaging kernel $\mathbf{A}^{1*}$. It is obvious that the processing ensures that the sensitivity for humidity (kernel block $\mathbf{A}^{1*}_{11}$) and for δD (kernel block $\mathbf{A}^{1*}_{22}$) are almost identical. Data with this characteristic are provided on a linear scale in the "ftir.iso.post.h2o" HDF files and they are well suited for $\{H_2O, \delta D\}$-pair distribution analyses.

4.4.2 A posteriori processing for a quasi-optimal estimation of $\{H_2O, \delta D, d\}$ triplets

Figure 6 reveals that kernel block $\mathbf{A}^{1*}_{33}$ is still rather different from the kernel blocks $\mathbf{A}^{1*}_{11}$ and $\mathbf{A}^{1*}_{22}$. Furthermore, there is a rather large cross dependency of δD on deuterium excess (kernel block $\mathbf{A}^{1*}_{32}$). This means that a posteriori correction with operator $\mathbf{C}$ according to Eq. (22) is not sufficient for providing deuterium excess data that can be analysed together with H_2O and δD and that is of sufficient quality. For such purpose, the a posteriori treatment has to be stronger, which can be achieved by using the following operator $\mathbf{C}$:

$$\mathbf{C} = \begin{pmatrix} \mathbf{A}^1_{33}{}' & 0 & 0 \\ -\mathbf{A}^1_{21}{}' & \mathbf{A}^1_{33}{}' & 0 \\ -\mathbf{A}^1_{31} & -\mathbf{A}^1_{32}{}' & \mathbb{1} \end{pmatrix}. \tag{23}$$

Figure 7 depicts the a posteriori processed averaging kernel $\mathbf{A}^{1*}$ by using $\mathbf{C}$ from Eq. (23). This treatment ensures that the three blocks $\mathbf{A}^{1*}_{11}$, $\mathbf{A}^{1*}_{22}$, and $\mathbf{A}^{1*}_{33}$ are almost identical. Nevertheless, the retrieved deuterium excess shows still some cross dependency on δD (averaging kernel block $\mathbf{A}^{1*}_{32}$). Currently, these data are not provided via the NDACC database, mainly because it has so far not been possible to empirically prove the quality of this deuterium-excess remote sensing data. In any case, interested users are welcome to investigate the quality of these data. The required a posteriori processing can be made by using the data provided in "ftir.iso.h2o" HDF files and by following the description given in this paper (if unclear, please contact the MUSICA team).

5 Data availability

The MUSICA NDACC/FTIR data are publicly available via the database of NDACC (ftp://ftp.cpc.ncep.noaa.gov/ndacc/MUSICA/) and via doi:10.5281/zenodo.48902 (Barthlott et al., 2016).
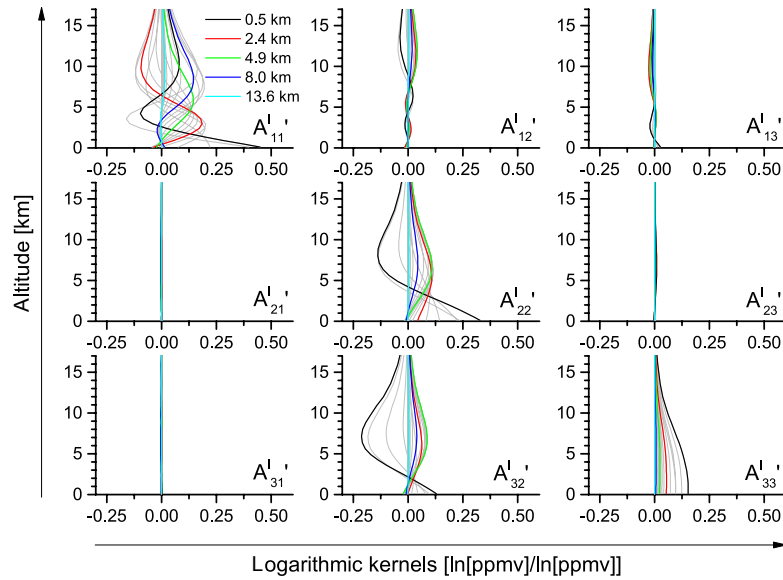
Figure 5. Column entries of the nine blocks of the logarithmic-scale kernel matrix in the {humidity, δD, d}-proxy state ($\mathbf{A}^{1'}$; see Eq. 18).
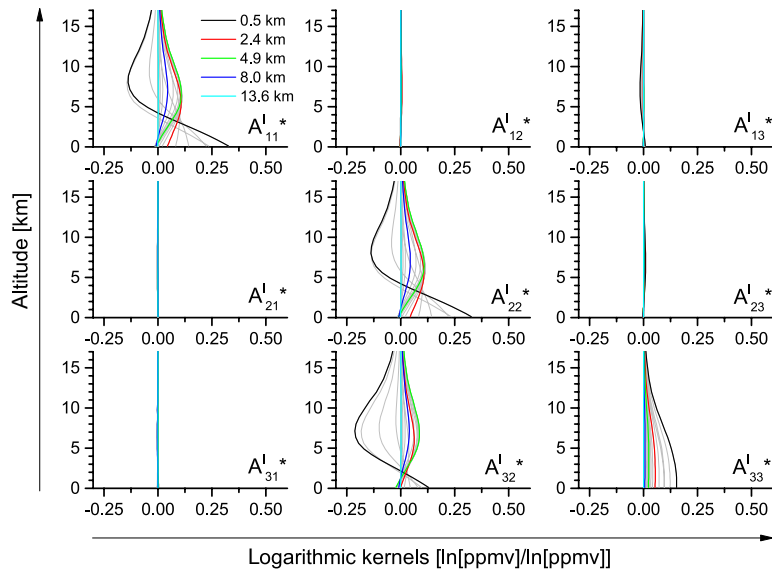

Figure 6. Same as Fig. 5 but after a posteriori processing with $\mathbf{C}$ from Eq. 22 ($\mathbf{A}^{1*}$; see Eq. 20).

6 Conclusions

For a correct optimal estimation retrieval of the full water vapour isotopologue state, we have to consider that atmospheric variations in different isotopologues are strongly correlated. This strong correlation is then also present in the retrieved state vectors and it has to be considered when interpreting averaging kernels and error covariances. As a consequence, it makes little sense to provide different isotopologues in the form of individual states and via individual data sets. Instead, it is essential that water vapour iso-

topologues are made available as single full state vectors together with their full averaging kernels and error covariances. The standard GEOMS metadata template for FTIR data on the NDACC database (called GEOMS-TE-FTIR) does not allow data to be provided in such a format and a slight extension of the standard FTIR template has been made (the modified template is called GEOMS-TE-FTIR-ISO-001). The MUSICA NDACC/FTIR data are now available in this new data format on the NDACC database and via doi:10.5281/zenodo.48902. The extended template can also
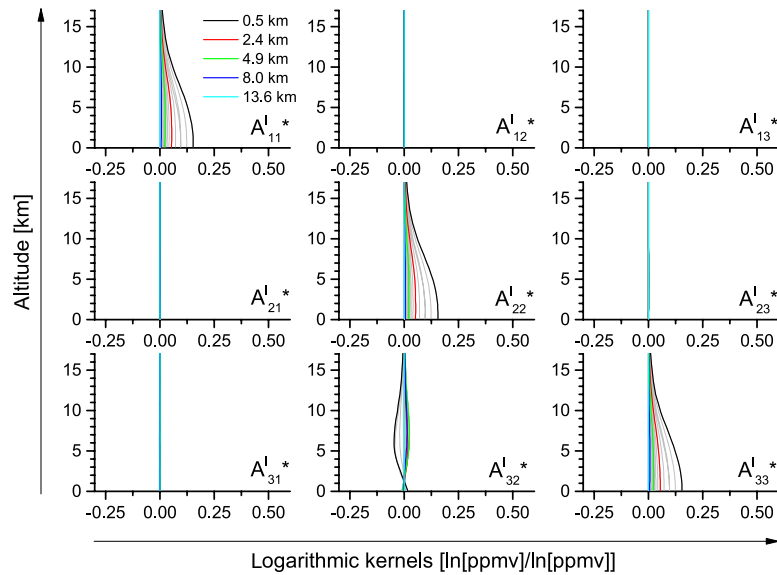
Figure 7. Same as Fig. 6 but using **C** from Eq. 23.

Table 3. Summary of recommendations and comments for the two principal types of data users.

User type	HDF file	Recommendations and comments
H_2O profiles	ftir.iso.h2o	transfer data onto a logarithmic scale (see Sect. 4.1) use $H_2^{16}O$ retrieval products as H_2O (more than 99.7 % of all water vapour is in form of $H_2^{16}O$) use averaging kernel $\mathbf{A}^1_{H_2O} = \mathbf{A}^1_{11} + \mathbf{A}^1_{12} + \mathbf{A}^1_{13}$ (see Sect. 4.3)
$\{H_2O, \delta D\}$ pairs	ftir.iso.post.h2o	comment: all "ftir.iso.post.h2o" retrieval products are corrected as described in Sect. 4.4.1 transfer data onto a logarithmic scale (see Sect. 4.1) use data in the {humidity, δD, d}-proxy basis for sensitivity and error assessments (see Sect. 4.4)

be used for providing isotopologue data of other molecules (e.g. ozone isotopologue data).

In order to be compliant with GEOMS, the data are provided on a linear scale and as volume mixing ratio (ppmv). However, since the retrieval is performed on a logarithmic scale, it is recommended to transfer states, kernels, and averaging kernels onto a logarithmic scale. On this scale, linearity in the context of averaging kernel operations can be assumed and we can furthermore make transformations between the $\left\{\ln\left[H_2^{16}O\right], \ln\left[H_2^{18}O\right], \ln\left[HD^{16}O\right]\right\}$ basis and the {humidity, δD, d}-proxy basis. The transformation of state vector, error covariances and averaging kernels onto the {humidity, δD, d}-proxy basis gives insight into the actual sensitivity and error characteristics of the retrieved humidity, δD, and deuterium-excess values. It can be shown that $\{H_2O, \delta D\}$-data pairs can be obtained in an optimal estimation sense, but only if the retrieval output undergoes a posteriori processing. Error estimations are discussed in detail in Schneider et al. (2012). The leading random error source is uncertainty in the atmospheric temperature profiles and artefacts in the spectral baseline (like channelling or offset). Sys-

tematic errors are dominated by uncertainties in the spectroscopic parameters. There are inconsistencies between parameters for different lines as well as inadequacies in the line shape model used. To determine the importance of the different aspects for improving the accuracy, further dedicated research projects would be needed.

The MUSICA NDACC/FTIR data are made available in the form of two different data types. The first type ("ftir.iso.h2o") is the direct retrieval output. It reports the optimal estimations of the states $\ln\left[H_2^{16}O\right]$, $\ln\left[H_2^{18}O\right]$, and $\ln\left[HD^{16}O\right]$, and a user should work with these data for studies that focus on water vapour and disregard the difference between the isotopologues. The second type ("ftir.iso.post.h2o") is the a posteriori processed output. It reports the optimal estimation of $\{H_2O, \delta D\}$ pairs and it should be used for analysing moisture pathways by means of $\{H_2O, \delta D\}$-distribution plots. Table 3 summarises the guidelines that should be considered for correct data usage.

Appendix A: Mathematical foundations of atmospheric remote sensing

Atmospheric remote sensing means that the atmospheric state is retrieved from the radiation measured after having interacted with the atmosphere. For a mathematical treatment, vectors are used for describing the atmospheric state and the measured radiation. Hence, vector and matrix algebra is the tool used in the context of remote sensing retrievals and remote sensing product characterisation. In the following, we briefly explain the connections between vector and matrix algebra and atmospheric remote sensing that are relevant for our paper. For a more detailed introduction, please refer to Rodgers (2000). For a general introduction to vector and matrix algebra, dedicated textbooks should be consulted (e.g. Herstein and Winter, 1988; Strang, 2016).

A1 Atmospheric state vector and measurement vector (x and y):

We are interested in the vertical distribution of $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ and our atmospheric state is a vector whose components represent the $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ concentrations at different atmospheric altitudes. The atmospheric state vector is generally written as x. Our measurements are high-resolution infrared spectra and the different spectral bins of a measured spectrum are represented by the components of the measurement vector. This vector is generally written as y.

A2 A priori state vector and covariance matrix (x_a and S_a)

The interaction of radiation with the atmosphere is modelled by a radiative transport model (also called a forward model, F) and allows relating the measurement vector and the atmospheric state vector by

$$y = F(x). \tag{A1}$$

We measure y and are interested in x. However, a direct inversion of Eq. (A1) is generally not possible, because there are many atmospheric states x that can explain one and the same measurement y. This means we face an ill-posed problem and for the inversion we need an additional constraint or regularisation. This leads us to the optimal estimation method. For this method we set up a cost function that combines the information provided by the measurement with a priori known characteristics of the atmospheric state (see Eq. 1 in Sect. 2.1). This a priori knowledge about the atmospheric state is collected as the mean and covariances in form of the a priori state vector (x_a) and the a priori covariance matrix (S_a). By minimising the cost function (Eq. 1), we get the most probable atmospheric state for the given measurement, i.e. the optimally estimated atmospheric state.

A3 Averaging kernel matrix (A)

The averaging kernel relates the real atmospheric state to the atmospheric state as provided by the remote sensing retrieval (see, for instance, Eq. 13). The averaging kernel matrix element (i, j, i.e. element in line i and column j of the matrix) reveals how a change in the real atmospheric state vector component j will affect the retrieved atmospheric state vector component i. The averaging kernel is an important component of a remote sensing retrieval and it is calculated by the retrieval code by combining the forward model calculations with the retrieval constraints:

$$A = \left(K^T S_\epsilon^{-1} K + S_a^{-1} \right)^{-1} K^T S_\epsilon^{-1} K. \tag{A2}$$

K is the Jacobian matrix (derivatives that capture how the measurement vector will change for changes in the atmospheric state) and provided by the forward model. S_ϵ and S_a are measurement noise covariance matrix and a priori atmospheric state covariance matrix.

A4 Basis systems for describing the atmospheric state

A coordinate system for a vector space is called a basis and consists of linearly independent unit vectors. Vector spaces can be equivalently described by different bases. One possibility for describing the atmospheric water vapour isotopologue state is to use three independent basis systems for the $H_2^{16}O$, $H_2^{18}O$, and $HD^{16}O$ state vectors. This is what we call in the paper the $\{ \ln[H_2^{16}O], \ln[H_2^{18}O], \ln[HD^{16}O] \}$ basis. Another possibility is to describe the atmospheric water vapour isotopologue state by means of three basis systems, whose respective vector spaces vary almost independently in the real atmosphere. This is what we call in the paper the $\{\text{humidity}, \delta D, d\}$-proxy basis. The transformation between the two different basis systems is achieved by the transformation operator P (see Eq. 5).

Both basis systems are equivalent, but using both of them helps to correctly constrain the inversion problem and to adequately describe the characteristics of the remote sensing product. The retrieval processor works in the $\{ \ln[H_2^{16}O], \ln[H_2^{18}O], \ln[HD^{16}O] \}$ basis and we need to formulate S_a in this basis according to Eq. (9). Furthermore, for the NDACC database we provide the data in this basis (or actually in the $\{ H_2^{16}O, H_2^{18}O, HD^{16}O \}$ basis; for transfer from a logarithmic basis to linear basis, see Sect. 4.1) in order to achieve compliance with standard metadata templates. However, it is the $\{\text{humidity}, \delta D, d\}$-proxy basis system in which the data product can be best described. That is the reason why in the paper we switch between the two basis systems and between logarithmic and linear scales.

Appendix B: Modification of HITRAN 2012 line parameter

For the ground-based FTIR retrieval, consideration of a non-Voigt line shape parameterisation becomes important because of the very high-resolution spectra (full width at half maximum of the instrumental line shape of about $0.005\,\mathrm{cm}^{-1}$).

We allow for a speed-dependent Voigt line shape, and in doing so we assume a Γ_2/Γ_0 of 15 %, which is in good agreement with previous studies (e.g. D'Eu et al., 2002; Schneider et al., 2011) and fit line intensity (S) and pressure broadening (γ_{air}). We use six high-quality H_2O and δD in situ profiles measured during an aircraft campaign in the surroundings of Tenerife and in coincidence with ground-based FTIR observations (Dyroff et al., 2015; Schneider et al., 2015) for empirically estimating the overall errors in S and γ_{air}. In addition, we use FTIR spectra measured at three distinct sites (Izaña, Karlsruhe, and Kiruna) for eliminating inconsistencies between the parameters of the different lines. Theory, practice, and limitations of such empirical line parameter optimisation method are discussed in Schneider and Hase (2009) and Schneider et al. (2011).

Table B1 resumes the modifications we had to make on the HITRAN 2012 parameters for different lines of Fig. 1 in order to adjust them for a speed-dependent Voigt line shape and for bringing them into agreement with coincident ISOWAT profile measurements and for minimising the residuals in the spectral fits at Izaña, Karlsruhe, and Kiruna. The obtained values are in agreement with our previous studies (Schneider and Hase, 2009; Schneider et al., 2011) and they are reasonable in the sense that they lie within the uncertainty ranges as given in the HITRAN data files. A value for Γ_2/Γ_0 of 15 % means a line narrowing, which in a Voigt line shape model could be approximated by reducing γ_{air} by 4 %. In order to counterbalance, parameter γ_{air} had to be generally increased (see last column in Table B1).

Table B1. Modifications in the line parameters (line intensity and pressure broadening) made with respect to HITRAN 2012.

Line centre (cm^{-1})	Isotopologue	ΔS (%)	$\Delta \gamma$ (%)
2660.511700	$HD^{16}O$	-5.52	$+3.96$
2663.285820	$HD^{16}O$	-5.53	$+4.00$
2713.862650	$HD^{16}O$	-5.53	$+4.07$
2732.493160	$H_2^{16}O$	$+12.26$	$+9.35$
2819.449040	$H_2^{16}O$	-3.07	$+4.52$
2879.706660	$H_2^{16}O$	-8.26	$+6.84$
2893.075920	$H_2^{16}O$	-9.07	$+9.64$
3019.824500	$H_2^{18}O$	-5.40	-0.72
3052.444870	$H_2^{18}O$	-6.32	-0.71

www.earth-syst-sci-data.net/9/15/2017/

Earth Syst. Sci. Data, 9, 15–29, 2017

References

Aemisegger, F., Pfahl, S., Sodemann, H., Lehner, I., Seneviratne, S. I., and Wernli, H.: Deuterium excess as a proxy for continental moisture recycling and plant transpiration, Atmos. Chem. Phys., 14, 4029–4054, doi:10.5194/acp-14-4029-2014, 2014.

Barthlott, S., Schneider, M., Hase, F., Wiegele, A., Christner, E., González, Y., Blumenstock, T., Dohe, S., García, O. E., Sepúlveda, E., Strong, K., Mendonca, J., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Mahieu, E., Jones, N., Griffith, D. W. T., Velazco, V. A., Smale, D., Robinson, J., Kivi, R., Heikkinen, P., and Raffalski, U.: Using XCO_2 retrievals for assessing the long-term consistency of NDACC/FTIR data sets, Atmos. Meas. Tech., 8, 1555–1573, doi:10.5194/amt-8-1555-2015, 2015.

Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Mengistu Tsidu, G., Grutter de la Mora, M., Strong, K., Notholt, J., Mahieu, E., Jones, N., and Smale, D.: The ground-based MUSICA dataset: Tropospheric water vapour isotopologues (H216O, H218O and HD16O) as obtained from NDACC/FTIR solar absorption spectra, Zenodo, doi:10.5281/zenodo.48902, 2016.

Craig, H.: Isotopic variations in meteoric waters, Science, 133, 1702–1703, 1961a.

Craig, H.: Standard for Reporting concentrations of Deuterium and Oxygen-18 in natural waters, Science, 13, 1833–1834, doi:10.1126/science.133.3467.1833, 1961b.

Craig, H. and Gordon, L. I.: Deuterium and oxygen 18 variations in the ocean and marine atmosphere, in: Stable Isotopes in Oceanographic Studies and Paleotemperatures, 1965, Spoleto, Italy, edited by: Tongiogi, E., 9–130, V. Lishi e F., Pisa, 1965.

Dansgaard, W.: Stable isotopes in precipitation, Tellus A, 16, 436–468, doi:10.1111/j.2153-3490.1964.tb00181.x, 1964.

D'Eu, J.-F., Lemoine, B., and Rohart, F.: Infrared HCN Lineshapes as a Test of Galatry and Speed-dependent Voigt Profiles, J. Mol. Spectrosc., 212, 96–110, 2002.

Dyroff, C., Sanati, S., Christner, E., Zahn, A., Balzer, M., Bouquet, H., McManus, J. B., González-Ramos, Y., and Schneider, M.: Airborne in situ vertical profiling of HDO / $H_2^{16}O$ in the subtropical troposphere during the MUSICA remote sensing validation campaign, Atmos. Meas. Tech., 8, 2037–2049, doi:10.5194/amt-8-2037-2015, 2015.

Galewsky, J., Sobel, A., and Held, I.: Diagnosis of Subtropical Humidity Dynamics Using Tracers of Last Saturation, J. Atmos. Sci., 62, 3353–3367, doi:10.1175/jas3533.1, 2005.

Gat, J. R.: Atmospheric water balance – the isotopic perspective, Hydrol. Process., 14, 1357–1369, doi:10.1002/1099-1085(20000615)14:8<1357::AID-HYP986>3.0.CO;2-7, 2000.

González, Y., Schneider, M., Dyroff, C., Rodríguez, S., Christner, E., García, O. E., Cuevas, E., Bustos, J. J., Ramos, R., Guirado-Fuentes, C., Barthlott, S., Wiegele, A., and Sepúlveda, E.: Detecting moisture transport pathways to the subtropical North Atlantic free troposphere using paired H_2O-δD in situ measure-

ments, Atmos. Chem. Phys., 16, 4251–4269, doi:10.5194/acp-16-4251-2016, 2016.

Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S.: Intercomparison of retrieval codes used for the analysis of high-resolution, J. Quant. Spectrosc. Ra., 87, 25–52, 2004.

Herstein, I. N. and Winter, D. J.: Matrix theory and linear algebra, Macmillan, Collier Macmillan, 1988.

Noone, D.: Pairing Measurements of the Water Vapor Isotope Ratio with Humidity to Deduce Atmospheric Moistening and Dehydration in the Tropical Midtroposphere, J. Climate, 25, 4476–4494, doi:10.1175/JCLI-D-11-00582.1, 2012.

Pfahl, S. and Sodemann, H.: What controls deuterium excess in global precipitation?, Clim. Past, 10, 771–781, doi:10.5194/cp-10-771-2014, 2014.

Rodgers, C. D.: Inverse methods for atmospheric sounding – Theory and practice, Ser. on Atmos. Oceanic and Planet. Phys., World Sci., Singapore, doi:10.1142/9789812813718, 2000.

Rothman, L. S., Gordon, I. E., Barbe, A., Chris Benner, D., Bernath, P. F., Birk, M., Boudon, V., Brown, L. R., Campargue, A., Champion, J.-P., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Fally, S., Flaud, J.-M., Gamache, R. R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W. J., Mandin, J.-Y., Massie, S. T., Mikhailenko, S. N., Miller, C. E., Moazzen-Ahmadi, N., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V. I., Perrin, A., Predoi-Cross, A., Rinsland, C. P., Rotger, M., Simecková, M., Smith, M. A. H., Sung, K., Tashkun, S. A., Tennyson, J., Toth, R. A., Vandaele, A. C., and Vander-Auwera, J.: The HITRAN 2008 molecular spectroscopic database, J. Quant. Spectrosc. Ra., 110, 533–572, doi:10.1016/j.jqsrt.2009.02.013, 2009.

Rothman, L. S., Gordon, I., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P., Birk, M., Bizzocchi, L., Boudon, V., Brown, L., Campargue, A., Chance, K., Cohen, E., Coudert, L., Devi, V., Drouin, B., Fayt, A., Flaud, J.-M., Gamache, R., Harrison, J., Hartmann, J.-M., Hill, C., Hodges, J., Jacquemart, D., Jolly, A., Lamouroux, J., Roy, R. L., Li, G., Long, D., Lyulin, O., Mackie, C., Massie, S., Mikhailenko, S., Müller, H., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E., Richard, C., Smith, M., Starikova, E., Sung, K., Tashkun, S., Tennyson, J., Toon, G., Tyuterev, V., and Wagner, G.: The HITRAN2012 molecular spectroscopic database, J. Quant. Spectrosc. Ra., 130, 4–50, doi:10.1016/j.jqsrt.2013.07.002, 2013.

Schneider, M. and Hase, F.: Improving spectroscopic line parameters by means of atmospheric spectra: Theory and example for water vapour and solar absorption spectra, J. Quant. Spectrosc. Ra., 110, 1825–1839, doi:10.1016/j.jqsrt.2009.04.011, 2009.

Schneider, M., Hase, F., and Blumenstock, T.: Water vapour profiles by ground-based FTIR spectroscopy: study for an optimised retrieval and its validation, Atmos. Chem. Phys., 6, 811–830, doi:10.5194/acp-6-811-2006, 2006.

Schneider, M., Hase, F., Blavier, J.-F., Toon, G. C., and Leblanc, T.: An empirical study on the importance of a speed-dependent Voigt line shape model for tropospheric water vapor profile remote sensing, J. Quant. Spectrosc. Ra., 112, 465–474, doi:10.1016/j.jqsrt.2010.09.008, 2011.

Schneider, M., Barthlott, S., Hase, F., González, Y., Yoshimura, K., García, O. E., Sepúlveda, E., Gomez-Pelaez, A., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Wiegele, A., Christner, E., Strong, K., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Demoulin, P., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Ground-based remote sensing of tropospheric water vapour isotopologues within the project MUSICA, Atmos. Meas. Tech., 5, 3007–3027, doi:10.5194/amt-5-3007-2012, 2012.

Schneider, M., González, Y., Dyroff, C., Christner, E., Wiegele, A., Barthlott, S., García, O. E., Sepúlveda, E., Hase, F., Andrey, J., Blumenstock, T., Guirado, C., Ramos, R., and Rodríguez, S.: Empirical validation and proof of added value of MUSICA's tropospheric δD remote sensing products, Atmos. Meas. Tech., 8, 483–503, doi:10.5194/amt-8-483-2015, 2015.

Schneider, M., Wiegele, A., Barthlott, S., González, Y., Christner, E., Dyroff, C., García, O. E., Hase, F., Blumenstock, T., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Rodríguez, S., and Andrey, J.: Accomplishments of the MUSICA project to provide accurate, long-term, global and high-resolution observations of tropospheric {H_2O, δD} pairs – a review, Atmos. Meas. Tech., 9, 2845–2875, doi:10.5194/amt-9-2845-2016, 2016.

Steen-Larsen, H. C., Sveinbjörnsdottir, A. E., Peters, A. J., Masson-Delmotte, V., Guishard, M. P., Hsiao, G., Jouzel, J., Noone, D., Warren, J. K., and White, J. W. C.: Climatic controls on water vapor deuterium excess in the marine boundary layer of the North Atlantic based on 500 days of in situ, continuous measurements, Atmos. Chem. Phys., 14, 7741–7756, doi:10.5194/acp-14-7741-2014, 2014.

Strang, G.: Introduction to linear algebra, 5th Edn., Wellesley-Cambridge Press, 2016.

Yoshimura, K., Oki, T., and Ichiyanagi, K.: Evaluation of two-dimensional atmospheric water circulation fields in reanalyses by using precipitation isotopes databases, J. Geophys. Res., 109, D20109, doi:10.1029/2004JD004764, 2004.